MODELS OF WORKING MEMORY

Mechanisms of Active Maintenance and Executive Control

Edited by

AKIRA MIYAKE AND PRITI SHAH

1999

Chapter 3

CAMBRIDGE
UNIVERSITY PRESS

3 An Embedded-Processes Model of Working Memory

NELSON COWAN

FIVE CENTRAL FEATURES OF THE APPROACH

The embedded-processes model of working memory relies upon the following five principles, which emphasize links between memory and attention.

(1) Working memory information comes from hierarchically arranged faculties comprising: (a) long-term memory, (b) the subset of long-term memory that is currently activated, and (c) the subset of activated memory that is in the focus of attention and awareness.

(2) Different processing limits apply to different faculties. The focus of attention is basically capacity limited, whereas activation is time limited. The various limits are especially important under nonoptimal conditions, such as interference between items with similar features.

(3) The focus of attention is controlled conjointly by voluntary processes (a central executive system) and involuntary processes (the attentional orienting system).

(4) Stimuli with physical features that have remained relatively unchanged over time and are of no key importance to the individual still activate some features in memory, but they do not elicit awareness (i.e., there is habituation of orienting).

(5) Awareness influences processing. In perception it increases the number of features encoded, and in memory it allows new episodic representations to be available for explicit recall.

Two prior integrative reviews of information processing, an article (Cowan, 1988) and a book (Cowan, 1995), describe a view that will serve as my basis for discussing working memory. According to this view, working memory refers to cognitive processes that retain information in an unusually accessible state, suitable for carrying out any task with a mental component. The task may be language comprehension or production, problem solving, decision making, or other thought. Most such tasks require that certain information be kept in mind. For example, in language comprehension, if the first word is totally forgotten by the time the second or third word is perceived, one is in bad shape. The mnemonic functions preserving information that can be used to do the necessary work collectively make up working memory. This is a functional definition in that any processing mechanisms contributing to the desired outcome, which is the temporary availability of information, are said to participate in the working memory system. In contrast, some researchers appear to prefer to define working memory according to the mechanisms themselves (e.g., Engle, Kane, & Tuholski, Chapter 4, this volume; Schneider, Chapter 10, this volume). Though my framework has much in common with those of these researchers, a functional definition of working memory seems more likely to encourage a consideration of diverse relevant mechanisms.

This work was supported by NIH grant R01 HD-21338.

The boundaries of this definition are fuzzy. If the process holding information to be used in a mental task also holds information irrelevant to the task (mistakenly or unavoidably), then this irrelevant information still might be said to be in working memory. The same is true if an entire process is invoked without the task being facilitated at all (e.g., if a subject tried to use verbal rehearsal to assist in the recollection of a meaningless shape, but to no avail).

Figure 3.1, reproduced from Cowan (1988), suggests the components of working memory that form an "embedded processes" model. The large rectangle represents all of the information in long-term memory. The jagged shape represents that subset of memory in a temporarily heightened state of activation. The small circle represents the information in the current focus of attention or conscious awareness. That information is assumed to be a subset of the activated information because it is presumably not possible to attend to information without activating it. However, the converse is not true. Many studies suggest that it is possible to activate information automatically, outside of the focus of attention and awareness (e.g., Balota, 1983; Marcel, 1983; Wood & Cowan, 1995), though it still is a matter of debate exactly how much information, and what types, can be activated automatically (Cowan, 1995; Holender, 1986). At least the information in the focus of attention, and possibly all of the activated information, can result in new links between concurrent or consecutive activated elements, forming new composites that are entered into long-term memory (e.g., in learning new words or remembering new episodes).

Some theories of working memory equate it to the focus of attention and awareness (similar to the "primary memory" concept of James, 1890) and some equate it to the sum of activated information (similar to the activated cell assembly notion of Hebb, 1949). Often, the distinction between activation and awareness is left unclear, but I argue that the distinction is important and that working memory must involve both, and some long-term memory information as well.

Cowan (1988, 1995), like James (1890), considered "attention" and "awareness" to be coextensive, at least within neurologically intact individuals. Attention was seen as an enhancement of the processing of some information to the exclusion of other, concurrently available information. This

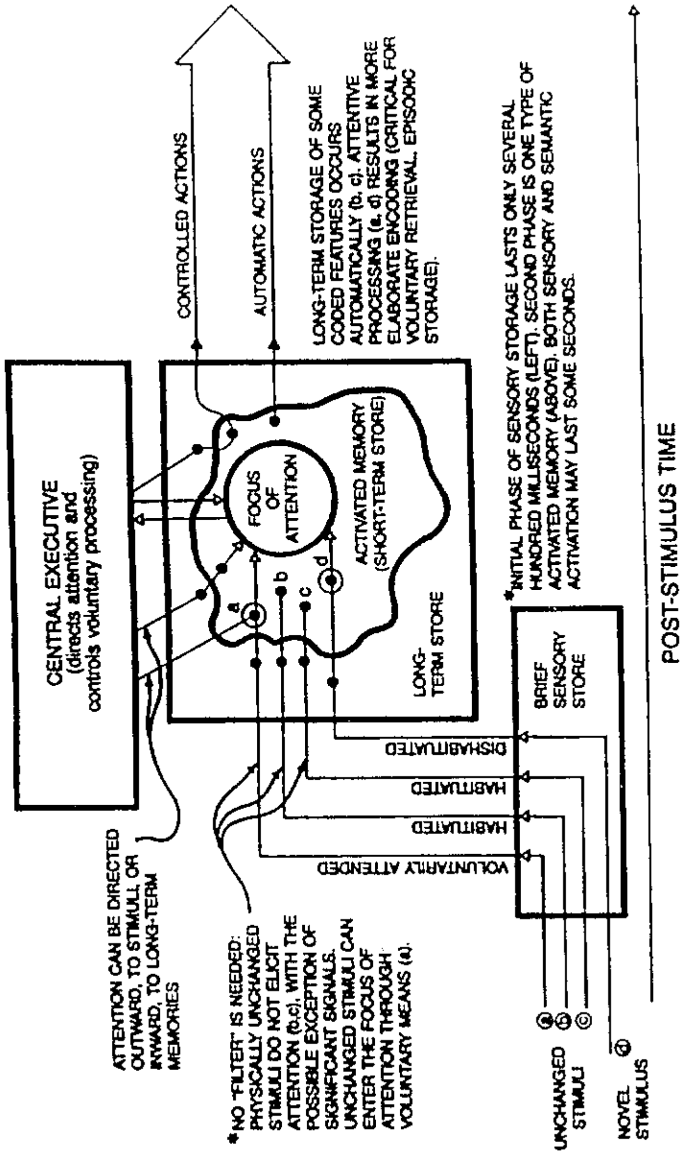


Figure 3.1. The Cowan (1988) model. Reprinted with permission of the American Psychological Association from Cowan (1988, p. 180, Figure 1). Copyright © 1988 by the American Psychological Association.

effect of attention was viewed as cutting across processing domains and tasks. For example, switching lanes on a highway probably is attention demanding in that it restricts diverse types of information processing, such as those involved in conversation or ongoing thought. In contrast, navigating the vehicle automatically according to well-learned geographical cues probably would not be considered attention demanding, even if it resulted in the failure to use other potentially available visual cues, unless it more generally restricted the ongoing stream of thought and voluntary actions. In this conception of attention, processing termed "automatic attending" (Shiffrin & Schneider, 1977), the automatically preferred processing of cues that have been well-learned, would not be considered attention demanding.

Although some activation of memory occurs for all stimuli, the sets of features of items entering the focus of attention are activated more completely than the sets corresponding to items out of the focus. An important item presented in an ignored channel while awareness is directed elsewhere, such as one's own name in an unattended voice, activates some features of the memory representation automatically. If these activated features are sufficient, they recruit attention to this channel, which then results in a more complete encoding. As shown in Figure 3.1, it is assumed that the allocation of attention is controlled jointly by (a) the automatic recruitment of attention to especially noticeable events (e.g., loud noises or bright lights, changes in physical properties of background stimulation, and sometimes signals that are of especially high personal relevance) and (b) voluntary, effort-demanding processes directed by the central executive (defined operationally as the collection of mental processes that can be modified by instructions or incentives).

Finally, it may be possible for attention to be directed away from particular features in memory rather than toward them. This inhibition of information in memory is something that many investigators have discussed lately (e.g., Conway & Engle, 1994; Gernsbacher, 1993; Hasher & Zacks, 1988). When I refer to activation and attention, I mean relative amounts, presumably including below-baseline amounts of activation for inhibited stimuli. Engle, Conway, Tuholski, and Shisler (1995) showed that inhibition appears to use the same resources as attentive processing in working memory, by demonstrating that negative priming was eliminated by a sufficiently large memory load (3 to 4 words, but not 1 to 2 words).

All three of the memory components (activation, the focus of attention and awareness, and long-term memory) contribute to working memory. The information in the current focus is the most readily accessible information in working memory. Information that is activated, but not to the point of conscious awareness, also can be retrieved reliably, albeit with slightly more of a delay. For example, when a subject searches through recently presented items to determine if a probe item was present among them (Sternberg, 1966), it is presumably this activated memory that must be searched. Finally, an item that is not even in an activated state still can be thought of as in working memory according to the foregoing definition, if there are cues in working memory that point to the item and raise the likelihood that it could be retrieved if necessary. This point will be addressed in the section on the relation of working memory to long-term memory (see also Cowan, 1995; Ericsson & Delaney, Chapter 8, this volume).

The theoretical model depicted in Figure 3.1 is not intended as a detailed description of processing that can be used to predict performance quantitatively. That would require many arbitrary assumptions to which I do not wish to be committed at present. Instead, the model is intended as a simple summary and organization of what I actually believe about attention and memory. It does incorporate a few common, key assumptions that I see as high priorities for research (e.g., assumptions that temporary memory activation outside of awareness exists, that the habituation of orienting serves as an attentional filter, and that there is a central attentional resource).

In the model there also is an attempt to reach a very general level of analysis. In doing so, many potential distinctions between processes are not discussed, but they are not denied either. For example, verbal and pictorial codes are very different and may be represented in separate parts of the brain and differing properties (e.g., the suitability for temporal versus spatial recall), but both are considered to be portions of long-term memory that can be activated in an analogous manner and that may well have similar temporal parameters of activation and decay. Covert verbal articulation and mental imagery are two processes that are very different, but both can be represented as learned routines that can be elicited through commands issued by the central executive to assist in a task. It is my belief that there is a gain in clarity from sometimes ignoring differences between particular materials, modalities, and codes to think about the system as a whole. Also, we could not be sure at this point about which domain distinctions would be most fundamental.

My tentative answers to the designated questions (summarized in Table 3.1) will be illustrated with results from my own recent research, to clarify how one might address the questions from this viewpoint and where it leads.

Basic Mechanisms and Representations in Working Memory

These include mechanisms of encoding, representation, maintenance, and retrieval of information, to be addressed in turn.

Encoding

A stimulus presumably activates features of memory, and the composite of activated features forms the encoding of the stimulus used in working memory. However, the activation of features is only partial if the stimulus is unattended to when it is presented. Under these circumstances, physical features are more likely to be represented than are semantic features. If the stimulus

(or its sensory afterimage) are attended to, then more features corresponding to the stimulus will be activated. A more extensive activation of features also results in a more stable memory representation.

The emphasis on multiple codes for a stimulus seems present also in Baddeley and Logie (Chapter 2, this volume), Engle et al. (Chapter 4, this volume), and Schneider (Chapter 10, this volume). However, those investigators stress abstract codes such as the phonological code (derived from either spoken or visual language stimuli) and the semantic code (derived from any potentially meaningful stimulus). I emphasize the importance of these codes also, but I may place more emphasis on sensory codes available only for a specific modality of input. One must consider all available codes to understand how items are preserved in working memory.

A study by Cowan, Lichty, and Grove (1990) illustrates the generation of multiple codes. They examined the memory of consonant–vowel syllables (*bee, dee, gee, bih, dih, gih, beh, deh,* and *geh*) that were presented one at a time, at irregular intervals of 2 to 13 s, and were to be ignored while the subject was busy silently reading a novel. Once in a while, the reading task would be interrupted by a recall signal light; then the subject was to put down the book

Table 3.1. *Embedded-Processes Model: Brief Summary of Answers to the Eight Designated Questions*

(1) Basic Mechanisms and Representations in Working Memory
The model (Figure 3.1) was described by Cowan (1988, 1995). Working memory is a complex construct involving all information accessed for a task, including (a) memory in the focus of attention, (b) memory out of the focus but nevertheless temporarily activated, and (c) inactive elements of memory with sufficiently pertinent retrieval cues. The organization is embedded, with active memory as a subset of long-term memory and the focus of attention as a subset of active memory.

(2) The Control and Regulation of Working Memory
Operationally defined, the "central executive" is the set of processes influenced by instructions or incentives. The direction of the attentional focus is controlled jointly by (a) the central executive and (b) automatic recruitment of attention to physically changed stimuli or, occasionally, stimuli with special significance to the subject (Wood & Cowan, 1995). When the stimuli are unchanged, habituation of the attentional orienting response occurs, making it easier for the central executive to control the attentional focus.

(3) The Unitary Versus Non-Unitary Nature of Working Memory
The concept of "activated memory" subsumes activation that is sensory and abstract and that is based on any modality and any form of representation. In this sense, the model is unitary. (It is acknowledged that forms of representa-

continued

Table 3.1, continued

tion have different consequences for processing, but they all may be included under the rule that interference occurs between representations that are similar.) The model is less unitary than models that do not include activation outside of attention.

(4) The Nature of Working Memory Limitations
Each aspect of working memory has some limit. The evidence from various types of stimulus and coding modalities suggests that there is a time limit in the activation of memory, with activation fading within about 10 to 20 s unless it is reactivated (Cowan et al., 1990, 1994, 1997). In contrast, the focus of attention is limited by its capacity rather than by time. It appears to be limited to very few (3 to 5) unrelated items, though chunking and structure can raise the effective limit.

(5) The Role of Working Memory in Complex Cognitive Activities
People will obtain the information needed for complex tasks from any source available. Thus, information will be held in the focus of attention when that is possible, and more information may be kept active in memory, outside of the focus, when the capacity of the focus is exceeded. For example, rehearsal may serve to recirculate items into the focus, reactivating them. Long-term memory may also be used if relevant information can be retrieved.

(6) The Relationship of Working Memory to Long-Term Memory and Knowledge
Information in long-term memory is activated when it is to be used in a task. If it isn't possible for all of the necessary information to be activated, additional long-term memory information sometimes can be retrieved as necessary. Often, though, for success in a task, certain pieces of information must be activated concurrently to be combined. New combinations of information can be formed within active memory, which then may become part of long-term memory.

(7) The Relationship of Working Memory to Attention and Consciousness
Working memory includes the focus of attention, which holds the information of which the person is conscious (in neurologically intact individuals, though attention and consciousness can be dissociated in split-brain patients). However, working memory also includes activated memory outside of attention or conscious awareness.

(8) The Biological Implementation of Working Memory
Cowan et al. (1993) demonstrated memory activation using mismatch negativity responses within auditory event-related potentials. The mismatch negativity was shown to result when a deviant tone differed from a repeating standard tone, but only if the standard's representation was active when the deviant was presented. Cowan (1995) suggested biological underpinnings of the major aspects of the model (see Figure 3.9). For example, the inferior parietal areas were suggested as critical in representing the focus of attention.

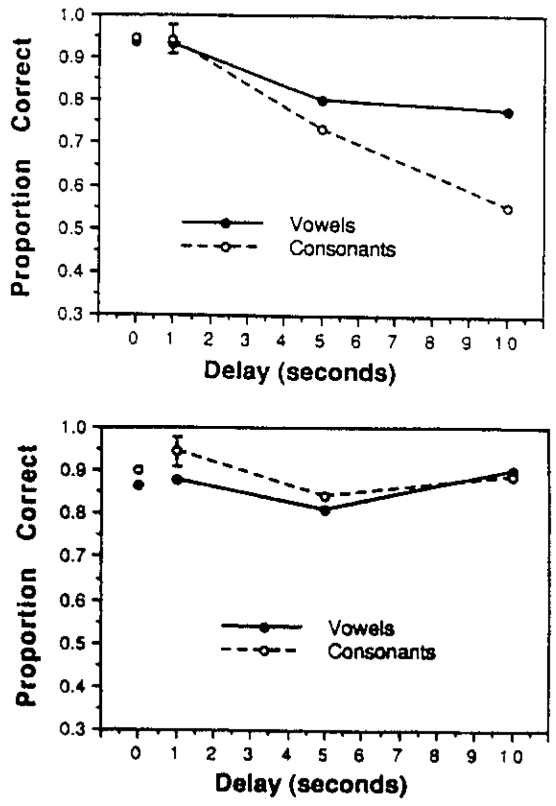


Figure 3.2. Memory for consonants and vowels within syllables presented in an auditory manner that was ignored (top panel) or monitored for occurrences of the syllable "dih" (bottom panel) during a silent reading task. Reprinted with permission of the American Psychological Association from Cowan, Lichty, and Grove (1990, p. 260, Figure 1, and p. 263, Figure 3). Copyright © 1990 by the American Psychological Association.

and identify the last syllable presented, and write a sentence about what was happening in the reading. The recall signal could come 1, 5, or 10 s after the last syllable. Finally, there was a comprehension test on the reading.

Figure 3.2 (top panel) shows the speech memory result, scored separately for the consonants and vowels. Consonant recall dropped quickly across retention intervals, but the acoustically simpler vowels were better preserved across retention intervals. Another experiment showed that this difference between consonants and vowels was not just a matter of phoneme order within a syllable; a consonant advantage was not obtained when the stimuli were vowel–consonant syllables. Reading comprehension in these experiments was no worse than for a control group that did not receive sounds while reading. The results suggest that speech information was activated automatically but that the acoustically more complex consonant information did not last as long in memory as the acoustically simpler vowel information.

In another experiment of Cowan et al. (1990), attention was divided between reading and listening. The stimuli were as in the first experiment, but the subject had the added task of pressing a button whenever the sound "dih" was heard. (Reading comprehension was hurt only slightly, but subjects found this dual task much more difficult than in the other experiments in the series.) The results under these divided attention conditions are as shown in Figure 3.2 (bottom panel). The consonant/vowel difference no longer appeared, and performance no longer was lost rapidly across retention intervals. Under these conditions, encoding was enhanced enough to produce a longer-lasting memory representation. If the speech representation that was formed with the help of attention was categorical rather than simply acoustical in nature, that could explain why performance in this experiment was equally good for consonants and vowels despite their difference in acoustic complexity. Thus, attention can change the nature of perceptual encoding dramatically.

In a final experiment of Cowan et al. (1990), subjects whispered the reading instead of reading silently, and the whispering response was recorded to allow an on-line indication of attention. A break in audible reading throughout either the 1-s period before or the 1-s period after the onset of the target syllable was considered as potential evidence of an attention shift toward that syllable. In 83% of the trials there was no such evidence of attention-shifting, and in those trials performance looked like the overall results. In the other 17% of the trials, in which there was a break in whispered reading at a critical time, the results were different. Consistent with the speculation that attention sometimes shifted toward the spoken syllable, the consonant–vowel difference was much smaller in these trials, with performance levels at a 1-s retention interval for the consonants (93% correct) being much higher than in the trials with no evidence of attention shifts (70% correct). These results reinforce the conclusion that attention enhances the encoding process and may be necessary for a categorical level of coding.

Some well-known research suggests that some semantic processing can occur automatically, but it is not clear to what extent. For example, Moray (1959) found that although people can fully attend to only one channel of information at a time, they notice their names spoken in an unattended channel when questioned about it several minutes later (Moray, 1959). However, Moray performed this experiment only on a preliminary basis and the actual finding was that 4 of 12 subjects noticed their names. Surprisingly, we found no replications of the effect in the literature. In response, Wood and Cowan (1995) replicated the effect with a better-controlled method and more subjects. The percentage of subjects who noticed their names (34%) was quite similar to what Moray found, and an on-line measure of attention shifting (errors and pauses in shadowing or repetition of information in an attended message) yielded no evidence that subjects who noticed their names had been shifting attention beforehand to the ear in which the name occurred. A dramatic shift of attention occurred afterward, as shown in Figure 3.3. This result suggests that semantic encoding of unattended information is limited in its extent. Otherwise, an item as salient as the subject's own name should have occurred more frequently. Articles proposing that a more complete semantic encoding takes place without awareness have been criticized by Cowan (1988, 1995), Wood, Stadler, and Cowan (1997), and Holender (1986), among others.

Representation

The basic principles of encoding, maintenance, and retrieval could be similar no matter what the form of the representation of an item. Although there is extensive evidence that phonological short-term retention is impeded by competing phonological activity (articulatory suppression) whereas visuospatial short-term retention is impeded by competing spatial activity (Baddeley, 1986; Baddeley & Logie, Chapter 2, this volume), these phenomena could be subsumed under the principle that representations in memory are degraded by similar representations. A peculiarity in Baddeley's (1986) formulation that Cowan (1988) aimed to address is that Baddeley focused on two possible storage types and generally neglected others, such as the storage of nonverbal sounds or tactile sensations. Also, if various types of storage all have similar properties, then these similarities may be important, not just the differences between them.

At least some aspects of representation do seem similar across modalities. Cowan (1988) pointed out that in each modality (including at least vision, audition, and tactile senses), the ability to compare two slightly different stimuli declines rapidly as a function of the time between the two stimuli across about 10 to 20 s. Also, in each modality, it seems that the greatest interference comes from additional similar stimuli in that modality. The same appears true for types of internal code or representation (e.g., spatial or phonological coding). Interference comes from similar coding of subsequent stimuli (e.g., see Nairne, 1990; Cowan & Saults, 1995). Thus, similar decay properties over time and modality- or code-specific interference appear to characterize various types of temporarily activated memory.

Within this view, Baddeley's (1986) phonological loop and visuospatial sketchpad are just two varieties of memory activation, along with the processes that can be used to reactivate this memory (e.g., rehearsal or visualization). These reactivation routines are initiated by the central executive, although they can become to some extent automatized. Thus, as children grow older the amount of attention they need to carry out rehearsal decreases (Guttentag, 1984).

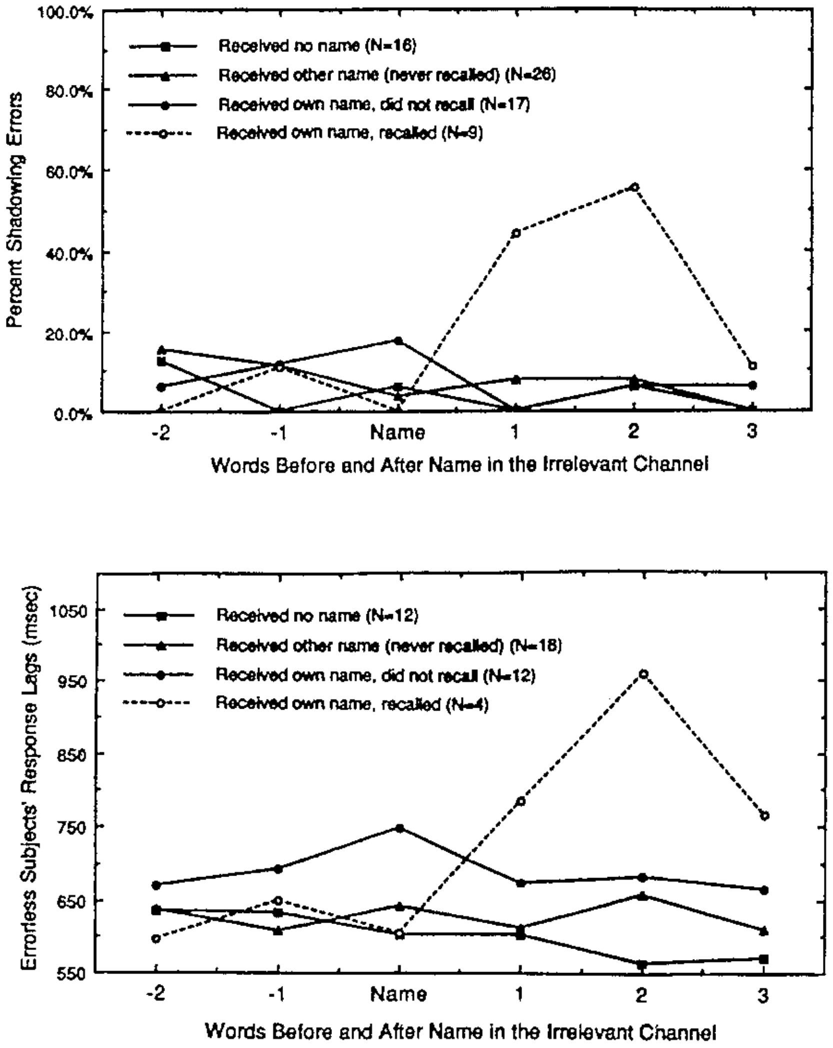


Figure 3.3. Performance in a shadowing task before and after the subject's name is presented in the ignored channel. Top panel, errors in shadowing; bottom panel, response lags within errorless responses. Reprinted with permission of the American Psychological Association from Wood and Cowan (1995, p. 258, Figure 1, and p. 259, Figure 2). Copyright © 1995 by the American Psychological Association.

The properties of the representation depend partly on the modality and physical properties of the stimulus and partly on the recoding of the stimulus. For example, phonological coding of visual materials (Conrad, 1964) presumably takes place because this coding is best suited to serial recall, whereas a visuospatial code seems best suited for following directions to a location. For a printed word, orthographic, phonological, and semantic codes may coexist in memory (Cowan & Saults, 1995).

The interference between different kinds of codes is limited. Take, for example, the interference between auditory sensory and auditory imagery codes. On the one hand, it is clear that people can use auditory imagery processes to construct sensory-like codes (Crowder, 1989). On the other hand, this auditory imagery is not interchangeable with a truly sensory memory record. Keller, Cowan, and Saults (1995) presented two tones to be compared, separated by a 0.5-s or a 10-s silent interval. During the longer intervals, subjects had to rehearse either a short tone series or short melody silently, or were permitted to spend the time rehearsing the first tone in the pair. Either type of filler task produced performance that was worse than in the rehearsal condition, as shown in Figure 3.4. Thus, auditory imagery constructed in the distracting tasks does not have exactly the same function as auditory sensory memory formed from an actual stimulus.

Maintenance

The form of maintenance depends on the aspect of working memory being considered and its limits that must be overcome. The most conventional meaning of maintenance is finding a way to continue *activating* items in memory. For example, Baddeley's (1986) model highlights verbal rehearsal as a means to reactivate items in a phonological store (see also Baddeley & Logie, Chapter 2, this volume). In the present view, keeping an item in the focus of attention would serve a similar function. Central executive processes may carry out particular operations that reactivate items in memory as a by-product. For example, Cowan (1992) proposed that the process of searching through a set of items can help to reactivate them. The items may be recirculated through the focus of attention.

These suggestions about the effect of a search on activation were inspired by measurements of the duration of elements within spoken responses in a memory-span task administered to 4-year-old children (Cowan, 1992). If all children have a similar persistence of information in memory (e.g., the approximately 2-s period proposed by Baddeley, 1986) and it cannot be altered, then that memory persistence should limit the potential duration of a list-recall response. Consequently, differences between individuals in the proportion recalled would be related to differences in the rate of pronunciation of the items within the fixed recall period. Contrary to this prediction, Cowan (1992) found that as the list length increased, the durations of silent periods between the words in the response increased markedly. (In contrast, the dura-

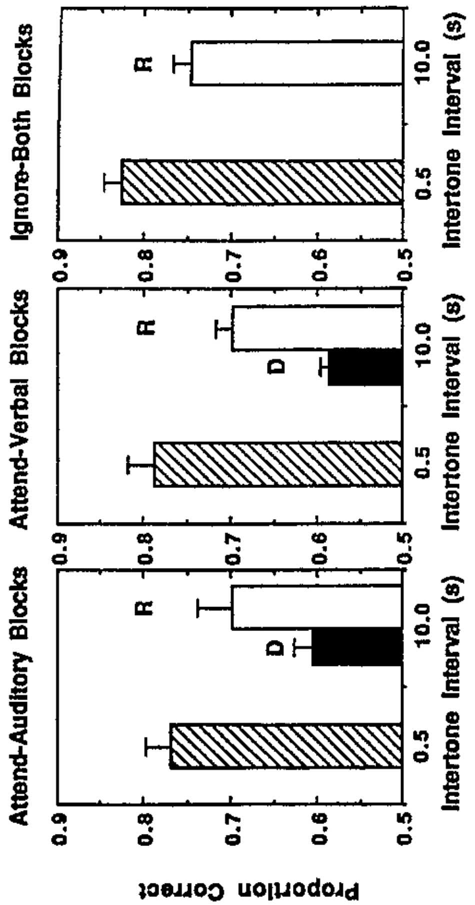


Figure 3.4. Tone comparison performance in several conditions. D = distraction during 10-s intertone intervals, R = rehearsal of the first tone during those intervals. Reprinted with permission of the American Psychological Association from Keller, Cowan, & Saults (1995, p. 642, Figure 3). Copyright © 1995 by the American Psychological Association.

tions of words did not change.) When one looked at the length of utterances in the span-length responses of various children, it turned out that the rates of production were similar; those who recalled more items did so in a response that lasted longer, with the duration of the recall response increasing by about 1 s for each word in the list. This suggested that the mental processing occurring during recall (primarily in the inter-item pauses) may have helped to reactivate the memories, allowing that longer response period.

The relevant activity that children carried out during inter-item pauses could be a mental search for the correct item to repeat next, rather than, say, verbal rehearsal. As each item is examined in the search, it may briefly enter the focus of attention and become reactivated. One clue that it is not a verbal rehearsal process taking place during these pauses is that they are unaffected by word length (Cowan, 1992; Cowan, Keller et al., 1994), which also is the case in memory scanning procedures (Chase, 1977; Clifton & Tash, 1973).

A subsequent study (Cowan et al., 1998) showed that subjects' ability to articulate quickly (estimating rehearsal speed?) and the durations of interword pauses in the memory span responses (estimating the speed of memory search processes?) did not correlate with each other at all, even though both of them produced moderate correlations with memory span (about .4). With multiple measures of each type of speed, the two types picked up different portions of the span variance and together accounted for a large portion of variance in a latent-variable model. The model explained about 60% of the total span variance and 87% of the age-related span variance on the basis of these two types of speed. Perhaps both speeds are important because memory decay can occur during either memory search or covert rehearsal.

Thus, various maintenance processes may work together or in sequence. In a trip to the store, one might encode the shopping list according to some meaningful scheme, rehearse the list several times, attend to key parts of it, mentally search for the item to acquire next, reconstruct part of the list from long-term memory, and so on. Recall ability may depend upon the success of the various processes.

Retrieval

Within the present model, retrieval means entering the correct items into the focus of attention. If an item is in the focus at the moment when it is needed, it probably can be recalled. Retrieval from activated memory and retrieval from long-term memory differ in important ways. Long-term memory has a richer information structure to rely upon. Also, retrieval from long-term memory is time limited only for practical reasons (e.g., the response period allowed on a recall trial in an experiment, or the amount of time one has to think of an acquaintance's name and introduce him or her to another person when it is expected socially). Retrieval from activated memory must occur quickly because the memory will not stay activated long. If the activated memory representation has disappeared, retrieval of the same information from long-term memory is still possible only if a sufficient episodic memory trace has been stored. Thus, retrieval processes must race against forgetting from activated memory. There also are tall hurdles in the race, in the form of interference among activated items.

Baddeley's (1986) working memory model is similar in principle to the current one, but the actual influences on retrieval seem more complex than he noted. In his description of memory span, a particular person can recall about as much from a particular stimulus set as he or she can pronounce in about 2 s. The theoretical explanation that Baddeley offered is that the speed at which one can pronounce the items estimates the rate of covert rehearsal. Both age and word length were said to affect rehearsal speed. Cowan, Keller, et al. (1994) found, however, that they work differently. The timing of spoken responses of children 4 and 8 years of age for lists of short (monosyllabic) or long (multisyllabic) words was measured. Better recall occurred for older children and shorter words, as expected. However, these variables resulted in different patterns of response timing. The word length (number of syllables) was found to alter the duration of words in the spoken responses, but not the duration of the silent periods. Conversely, the age of subjects altered the silent periods, but not the duration of words in the responses. Figure 3.5 shows aspects of the timing of correct responses to span-length lists. In preparatory intervals (top panel) there was only an effect of age, whereas in word durations (bottom panel) there was only an effect of stimulus word length. Interword pauses, not shown, revealed no effect in span-length lists, but they showed an age effect when examined for comparable list lengths at each age (mean = .38 s at age 4 versus .23 s at age 8). The response timing pattern suggests that the age and word length effects on span have different mechanisms. Both may arise only partly from covert rehearsal effects. The age effect may operate partly by influencing the speed of covert search processes during retrieval, whereas the word-length effect may operate partly by influencing the rate of overt pronunciation in recall (see Cowan et al., 1992). The efficiency of all such processes may contribute to span.

The Control and Regulation of Working Memory

In the present model, regulation of working memory amounts to the control of the focus of attention by the central executive. This control also is one way that information in memory can be activated; items in the focus become activated. However, the amount of information that can remain activated at one time is greater than the amount that would fit in the focus of attention (see also Engle et al., Chapter 4, this volume).

The model of Cowan (1988, 1995) has addressed the question of how voluntary and involuntary mechanisms work together to determine the focus of attention. The necessary background is as follows. Research on selective attention showed that it is easier to attend to one of several channels if they are distinguished by physical characteristics than if they are distinguished only by semantic characteristics (e.g., Broadbent, 1958; Johnston & Heinz, 1978). However, subsequent research showed that some of the semantic information is processed also (e.g., Gray & Wedderburn, 1960; Moray, 1959). Such findings led A. M. Treisman (1964) to adopt an "attenuating filter" theory of attention in which all information enters the processing system, but with attenuated

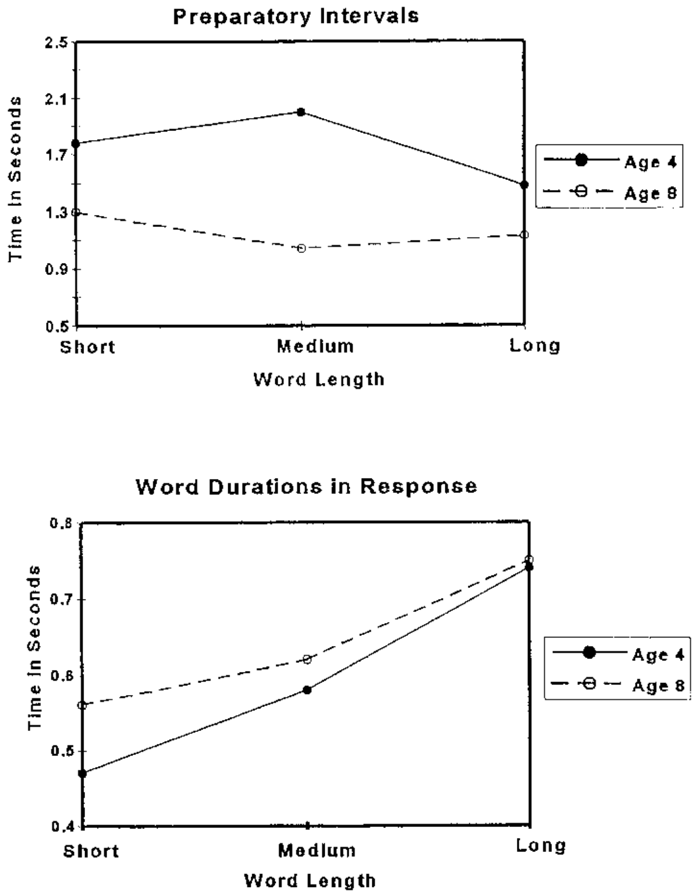


Figure 3.5. The timing of performance in correct responses to span-length lists in a word span task with short, medium, or long words. Top panel, preparatory intervals; bottom panel, word durations in the response. Adapted with permission of Academic Press, from Cowan, Keller et al. (1994). Interword pauses described in text but not shown in the figure.

processing of the unattended information and more complete processing of the attended information.

Cowan (1988, 1995) proposed a mechanism of control that serves a purpose similar to Treisman's (1964) notion of an attenuating filter, but with the mechanism of attenuation specified. Rather than unattended stimuli being filtered out, all stimuli activate some elements of memory, but this process is enhanced for attended stimuli (or for stimuli that recruit attention). The attention-recruiting mechanism is based on Sokolov's (1963) work on the orienting response and its habituation. The orienting response is a composite of physiological and behavioral responses, including temporarily reduced muscular activity, slowing of the heart rate, and increased sensitivity to incoming stimulation. It can be elicited when a new stimulus is presented, when the physical properties of a repeating stimulus are changed, or when a particularly significant stimulus is presented (Cowan, 1995; Öhman, 1979). However, when a stimulus is repeated with no change, habituation of the orienting response occurs. Sokolov's explanation was that a neural model of the stimulus is built up through repeated exposure. New stimuli are compared to the neural model, and an orienting response occurs only if there is a discrepancy.

The orienting response and its habituation are used in the model of Cowan (1988) to take the place of a selective filter or attenuator, as shown in Figure 3.1. All stimuli activate features in memory. However, attention is recruited to a stimulus through the orienting response only if the conditions are right. Therefore, if a subject wishes to ignore a repeated stimulus, he or she easily can do so; yet, some (mostly physical) features corresponding to the stimulus still will be activated in memory and compared to the neural model. Most complex semantic features of unattended stimuli will not be automatically processed and therefore are not available to be compared to the neural model (which itself may include attended features).

There are instances in which one needs to attend to a repeated stimulus. This can be done with input from the central executive, though only with effort. Thus, the focus of attention is controlled by the voluntary processes and automatic attentional recruitment together. Supporting this notion, Waters, McDonald, and Koresko (1977) had subjects carry out math problems, sometimes in the presence of auditory interference consisting of an audiotape of spoken mathematical terms ("plus," "equals," "three," etc.). Some of the subjects were allowed to hear the distracting audiotape beforehand, whereas others heard only control tones or nothing beforehand. Pre-exposure to the mathematical speech reduced the amount of distraction early in the session, as measured both by a reduction in errors and a physiological measure of the orienting response (skin resistance). There is also a recent finding of habituation and dishabituation to the disruptive effects of speech and office noise (Banbury & Berry, 1997).

In sum, the orienting mechanism and its habituation allow an effortless mode of attention in which some processing of the unattended information

takes place. The central executive uses effortful processes to help direct the control of attention. The effortless and effortful influences operate together. Presumably, if the physical properties of stimulation lead the subject toward the target object of attention (e.g., listening to an exciting lecturer in a quiet room) much less effort is needed for paying attention than when the physical properties lead away from the target object (e.g., listening to a lecture delivered in a monotonous voice over the cries of an infant in the room or a nearby siren).

Unitary Versus Non-Unitary Nature of Working Memory

The present model is unitary in some ways and non-unitary in others. The model seems more unitary than Baddeley's (1986) model, which makes a sharper distinction between phonological and visuospatial sources of information (see Baddeley & Logie, Chapter 2, this volume). Clearly, these types of information are quite different from one another, but different codes may be processed according to the same principles. It is not denied here that two items that require the same type of coding can interfere with one another more than two items that require different coding. However, coding distinctions other than phonological versus visuospatial may be equally important. The present view seems more unitary in this consideration. However, both models seem less unitary than that of Just and Carpenter (1992) or Lovett, Reder, and Lebiere (Chapter 5, this volume). Those models do not distinguish between activation and attention, which are discussed as if coextensive. Automatic activation apparently is outside the scope of their models. Like Baddeley's (1986) model, the present one postulates both a passive storage component and an active processing component. The modeling difference may hinge on whether one believes in the importance of automatic activation in working memory tasks.

It seems clear that all of the information reproduced in a simple memory span task could not result from the items in the focus of attention. It seems unlikely that, say, seven items could be held in attention at once. Therefore, in addition to attended information, one needs activated sources outside of attention and/or supplementary help from long-term memory. Thus, the present model includes several components and is not fully unitary.

The Nature of Working Memory Limitations

This question is probably critical in distinguishing between theories of working memory. In principle, there are at least two ways that a type of memory can be limited: in its capacity to hold information at any one time, and in the time for which information can be held. Both of these limitations are proposed, for different aspects of working memory (Cowan, 1995), with a time limit of memory activation and a capacity limit of the focus of attention. The time limit in activation is found also in the models of Baddeley (1986) and Baddeley and Logie (Chapter 2, this volume), Schneider (Chapter 10, this volume), and O'Reilly, Braver, and Cohen (Chapter 11, this volume), as well as in many older theoretical approaches (e.g., Broadbent, 1958), though I will argue that it is still difficult to obtain incontrovertible evidence for the time limit. The capacity limit of the focus of attention forms the basis of many approaches to working memory, including those of Engle et al. (Chapter 4, this volume) and Lovett et al. (Chapter 5, this volume), with some other researchers probably also being sympathetic (e.g., Young & Lewis, Chapter 7; Schneider, Chapter 10; O'Reilly et al., Chapter 11, this volume). It goes back to James (1890). Just and Carpenter (1992) use a similar limit but call it "activation," in contrast to the present terminology.

My approach appears to differ most from several investigators who seem more interested in constraints on processing arising from limitations in the available knowledge, its structure, or how it can be used (Young & Lewis, Chapter 7; Ericsson & Delaney, Chapter 8, this volume), or in how multiple tasks can be coordinated through the use of strategies (Kieras, Meyer, Mueller, & Seymour, Chapter 6; Barnard, Chapter 9, this volume). These additional limits are not denied here, either. The theories differ primarily in where one looks first for working memory limits in a particular task. Is the default hypothesis one of retention and capacity limitations, or one of strategic processing limitations? No one knows, but in my work (and below) the potential retention and capacity limitations are emphasized more.

Limitations of Activated Memory

The existence of limits on activated memory is critical for the present view. If there were no such limits, then activated memory would not really exist as a separate entity and could just as easily be conceived as simply the most easily retrievable items in memory, based on the current contextual cues.

CAPACITY LIMIT? Cowan (1988, 1995) suggested that the amount of activation that could take place at one time could be limitless. In contrast, Cantor and Engle (1993) argued that it is limited, based on their finding of larger fan effects in participants with a lower working memory span. They suggested that low-span participants have less activation to be shared among the learned examples, and therefore that their activation is spread thinner than high-span subjects when it must be shared among more than one item in an association fan. However, in a memory search task, Conway and Engle (1994) found that changes in response time as a function of set size were affected by working memory span only if each target item was used in sets of more than one size. The findings of Cantor and Engle were reinterpreted as resulting from high-span participants' greater ability to inhibit associated information that was irrelevant on a particular trial, not from the greater availability of activation in these participants. There is no known limit to the amount of activation, it appears.

TIME LIMIT? The other possible limitation of activated memory is that it decays. As time goes on, some of the activated features in memory cease to be activated, making the temporary representation of an item more and more vague until it disappears completely. Memory decay is a very common assumption. Baddeley (1986) assumed that activated elements in a phonological memory store decay in about 2 s, which is supposed to account for the fact that memory span is limited to the number of items that can be repeated in about that period. Many studies have been conducted in which two identical or slightly different tones or vowels are to be compared on each trial, and these studies have shown that performance declines as the time between sounds increases across about 10 to 30 s (see Cowan, 1984, 1995). Similar time estimates have come from studies of memory for several letters in the presence of a distracting task (e.g., Peterson & Peterson, 1959) and of the decline in the recency effect in list recall across a distracting period (e.g., Glanzer & Cunitz, 1966). The difference between these findings and the 2-s estimate of Baddeley (1986) could result from a number of differences among procedures in the required responses or sources of interference.

There has appeared, however, a body of research seriously challenging the notion that there is any kind of dependence on the passage of time per se. There was an article by Keppel and Underwood (1962) questioning the basis of forgetting in the procedure of Peterson and Peterson (1959). The latter found retention of consonant trigrams that declined steadily during a distracting task (counting backward by 3), with the performance reaching an asymptotically low level by 18 s. However, Keppel and Underwood found that this effect of the duration of distraction was not present in the first couple of trials, and emerged only after that. Their conclusion was that proactive interference from previous lists has to build up before the effect can be obtained.

Theoretically, there are two ways to interpret this finding of Keppel and Underwood (1962). One way is according to a theory in which there is no such thing as memory decay; there is a common or "unitary" set of memory principles that are supposed to apply to memory phenomena on any time scale (McGeoch, 1932; Crowder, 1993). According to this type of theory, in the Peterson and Peterson (1959) procedure the effect of the distracting task is to allow the temporal context to change. Shortly after the presentation of a trigram, it seems distinct from all of the others. However, after a long distractor task, the temporal distinctiveness of the final trigram is lost because the past few trigrams are now relatively close together compared to the retention interval.

According to an alternative interpretation, though, short- and long-term memory work together in this task (Cowan, 1988). The effect of the distracting task is presumably that it allows memory to decay. However, even when that decay has taken place, subjects will be able to retrieve the last trigram on the basis of a long-term memory representation, unless the proactive interference from previous trials is sufficient to prevent long-term retrieval. That presumably is what happens after a few trials.

A key claim of the unitary memory theory is that it is relative, not absolute, amounts of time that result in memory loss. More evidence for that claim came from a finding termed the "long-term recency effect." Prior studies had found that the recency effect (enhanced recall of items at the end of a list) was eliminated if a long (e.g., 20-s) distracting period was placed between the list and the recall period (e.g., Glanzer & Cunitz, 1966). However, in studies of the long-term recency effect (e.g., Bjork & Whitten, 1974), distracting periods are placed not only after the list, but also before the list and between items, a method usually termed the "continuous-distractor procedure." Under those circumstances, a recency effect is obtained even though the list-final distracting task should have been long enough to allow any activated memory to decay. The results follow a "ratio rule": performance is better when the ratio between the interpresentation interval and the retention interval is larger. The ratio is relatively large for the last few items in ordinary immediate recall and in the continuous-distractor procedure, but not for the last few items in delayed recall (uninterrupted presentation with a post-list distraction period). An explanation is that the recency effect reflects the special temporal distinctiveness of the last item or so when the time of recall is much closer to the end of the list than it is to previous list items, as is the case in the conditions in which the recency effect is obtained. Other research shows that several other effects usually taken as indicative of short-term memory also can be obtained in the continuous-distractor procedure (see Cowan, 1995, for a review).

A number of investigators have been looking for signs of similarities between recall in the ordinary procedure and the continuous-distractor procedure. However, it has been much less common to look for possible differences between these procedures, a strategy that potentially could reconfirm the concept of memory decay. Recently, we have pursued that strategy. Though several of the effects that had been taken as signatures of short-term memory and its decay have been compromised by the continuous-distractor results, one effect that has not been so compromised is the word-length effect. Cowan, Wood, and Borne (1994) examined this effect in the ordinary, immediate recall procedure and the continuous- (or "through-list") distractor procedure. Backward recall was used so that the words that could be rehearsed for the longest time were not the same as the ones to be recalled first, allowing rehearsal versus output delay mechanisms of the word-length effect to be distinguished. Word length was manipulated separately in the first and second halves of the list. The short words were monosyllabic, the long words trisyllabic.

The results are shown in Figure 3.6. In immediate recall, there was an advantage for lists in which the words to be recalled early on were short rather than long, replicating a previous finding (Cowan et al., 1992) and suggesting that items are forgotten during the time in which other items are recalled.

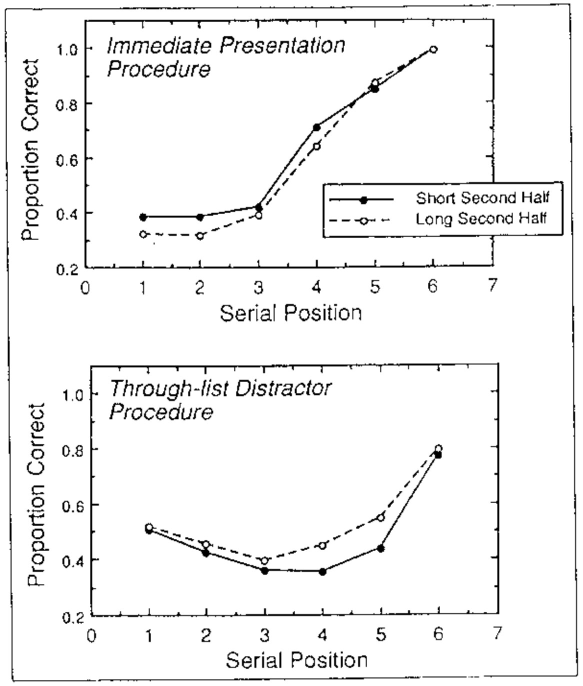


Figure 3.6. Performance on a backward span task for lists with short versus long words in the first half-list. Top panel, immediate recall; bottom panel, through-list distractor recall. Reprinted with permission of Cambridge University Press from Cowan, Wood, and Borne (1994, p. 105, Figure 1). Copyright © 1994 by Cambridge University Press.

However, in through-list (continuous) distractor recall, there actually was a long-word advantage, which might be attributed to the greater number of phonological cues in long-term memory. Thus, unlike other effects for which the immediate and continuous distractor procedures have been compared, the word length effect was not at all similar in the two procedures. Clearly, then, the word length effect may be particularly important in confirming the idea of the decay of activated memory.

However, the word length results have been criticized recently. The effect could result from the greater complexity, rather than the greater duration,of longer words. It is difficult to match sets of shorter and longer words on all potentially relevant phonemic features. On a suggestion by M. Treisman, an experiment was conducted by Cowan, Wood, Nugent, and Treisman (1997) to overcome this difficulty. One- and two-syllable, printed words were used, but the syllable number was not the only basis of word length. In addition, words were to be pronounced more quickly or slowly, as cued by a row of asterisks

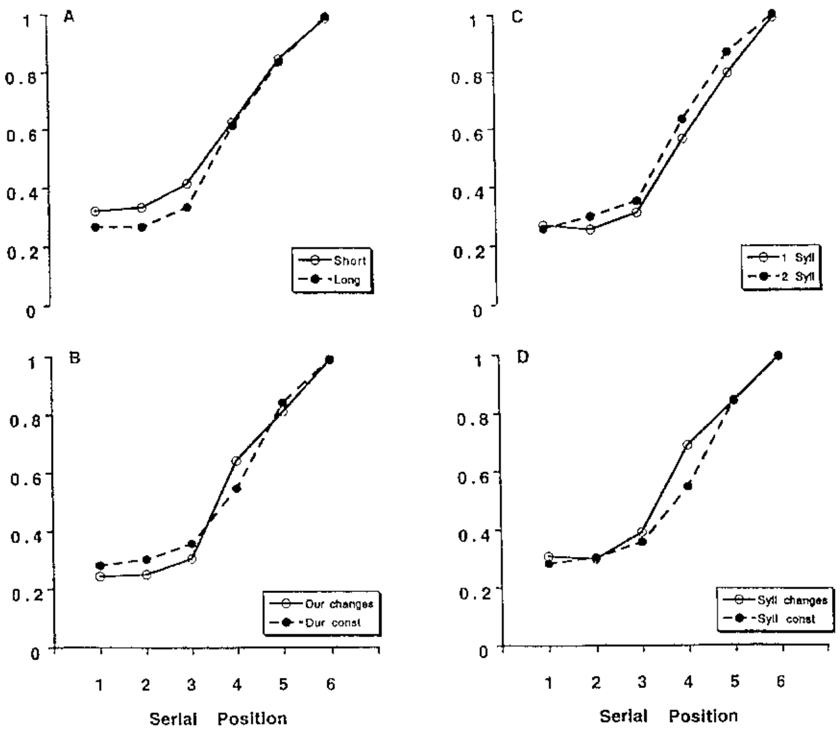


Figure 3.7. Performance in backward span for lists with words pronounced short versus long throughout (panel A), lists split according to word duration (panel B), lists with monosyllabic versus disyllabic words throughout (panel C), and lists split according to number of syllables (panel D). Reprinted with permission of Cambridge University Press from Cowan, Wood, Nugent, and Treisman (1997, Figure 2). Copyright © 1997 by Cambridge University Press.

that grew longer for either 300 ms or 600 ms. The same pronunciation duration cue appeared before a particular item at the time of presentation of the list and at the time of spoken recall. Recall was, once more, in the backward order. Figure 3.7 summarizes the most important results of this study. There was an advantage for words that were pronounced more quickly (panel A). The decay theory would predict this on the grounds that when words are pronounced more quickly, there is less time for forgetting to occur. There also was an unexpected advantage for disyllabic words over monosyllabic words, perhaps because the disyllabic words provide more phonological cues (panel C). This shows that word complexity is not the basis of the ordinary short-word advantage. The bottom panels (B and D) show that there were additional effects of splitting the list in terms of either duration or complexity.

As Treisman pointed out within our collaborative paper (Cowan et al., 1997), it still cannot be taken as proven that the word-length effect demonstrates the existence and importance of memory decay. The problem is that the longer-duration pronunciations expose subjects to their own speech for a longer time, and therefore the total amount of interfering content still is greater for words produced in the longer fashion. However, several previous investigators have obtained results with verbal stimuli that also seem to indicate the existence of verbal memory decay (Conrad & Hille, 1958; Reitman, 1974; Watkins, Watkins, Craik, & Mazuryk, 1973; Wingfield & Byrnes, 1972). Also, Cowan, Saults, and Nugent (1997) found separate roles of absolute and relative time in two-tone discriminations. Nevertheless, the concept of decay remains controversial in cognitive psychology (Crowder, 1993; Neath & Nairne, 1995) in light of potential nondecay accounts, and further evidence on the role of decay is still needed.

In sum, the notion of activation seems to depend on its temporal limitations, without which the activation would just appear to be a part of long-term memory with no special status except that it matches the current context well. This analysis of the meaning of the activation concept may explain, among other things, why decay plays a prominent role within the theory of Baddeley (1986). Although this is a difficult issue to examine, I have cited tentative evidence that memory activation is time limited.

Limitations in the Focus of Attention

CAPACITY LIMIT? Less controversial than the limits in activation discussed above, there is a limit to how many unrelated items can be in the focus of attention or awareness at any given time (James, 1890; Broadbent, 1958). It may be that people can focus on at most one general scheme (or several?) at a particular moment, not on many unrelated items or schemes.

A dissenting voice might be found in the late-filter theory of attention (e.g., Deutsch & Deutsch, 1963). According to that kind of view, the limit in attention is how much one can respond to at a particular time, not how much one can perceive or apprehend. Another dissenting view might be found in the multiple resources approach (e.g., Wickens, 1984), in which the idea of a central processing capacity limit is replaced with limits for specific types of material, with the total capacity depending on how dissimilar the streams of information are in various ways. However, it is not absolutely clear to me how, according to these alternative approaches, the behavioral aspects of attention and phenomenological aspects of awareness are supposed to be related to one another. Note that the suggestion that there is a central processing capacity (or attention) limit does not mean that there cannot also be more specific resource limits, such as a limit in verbal or spatial processing.

One of the simplest demonstrations of limits in attention comes from research on dichotic listening (e.g., Broadbent, 1958). Suppose that prose in a male voice is presented in one of the subject's ears and prose in a female voice is presented in the other ear. The subject will find it impossible to comprehend the meanings of both messages at once, even though both messages are acoustically clear. This is an easy demonstration to carry out. However, one might object that the limit could be in the language comprehension apparatus and not in the ability to hold multiple sources of information in mind once they have been comprehended.

A finding seemingly counter to the conclusion that attentional capacity is limited is that people can be trained to do two complex tasks at the same time (Hirst, Spelke, Reaves, Caharack, & Neisser, 1980). However, Cowan (1995) offered an extensive critique, suggesting that attention-switching or automaticity may well underlie such feats.

Most of the literature on capacity limits does require simultaneous perception of multiple channels, or multiple simultaneous responses. It is difficult to avoid the input and output requirements so as to examine the limits in what can be held in the focus of attention per se. Here one might think of limits in short-term memory (Miller, 1956) or in the size of the memory set within search tasks (Shiffrin & Schneider, 1977; Sternberg, 1966). However, these types of task might overestimate how much can be held in the focus of attention. In them, information might be kept active (e.g., through covert rehearsal) with only a small amount of the information actually residing in the focus of attention at one time. Nevertheless, these types of procedure at least do seem to indicate that there is some upper bound on the focus of attention.

There is, however, a seemingly "magical number" 4 ± 1. Derived measures of short-term memory in the immediate recall of verbal lists, designed to eliminate contributions of long-term memory, generally have suggested that no more than about 4 unconnected items can be kept in mind at one time (see Watkins, 1974). Converging evidence on this comes from a number of research domains. (a) A number of related phenomena were pointed out by Broadbent (1975), such as the limit in the number of items that occur in a cluster in free recall and the limit for obtaining perfect immediate memory performance. (b) On each trial within an experiment by Halford, Maybery, and Bain (1988) a memory list did or did not have marked similarities to the previous list. Proactive interference effects on search time were obtained with relatively long memory lists, but not with lists of 4 or fewer items. (c) In the classic research by Sperling (1960), a partial report cue was used to indicate to subjects which row of a briefly viewed, multicharacter array to report. When the partial report cue delay was short, subjects seemed to have access to most or all of the items in the array, which presumably were saved temporarily in an unanalyzed sensory form. However, with or without the cue, there also was a limit in the absolute number of items that could be reported on a particular trial. At most about 4 items could be reported, the "whole report" limit. One explanation for this limit is that only about 4 items can be drawn from sensory memory to the focus of awareness on a particular trial. A similar limit on whole report was

observed using spatiotemporal arrays in the auditory modality (Darwin, Turvey, & Crowder, 1972). (d) In research on subitizing (e.g., Trick & Pylyshyn, 1993) subjects are to count sets of small items such as dots. With 1 to 4 items, normal adult response times increase at the rate of about 40 to 120 ms per additional item, whereas starting with the fifth item, response times increase at the rate of 250 to 350 ms per item. This discontinuity has led to the notion that a "subitizing" or apprehension of the first 4 items takes place in parallel, with a deliberate, serial counting process necessary only with numbers larger than about 4. It could be argued that the subitized items are in the focus of attention at the same time. (e) In a recently developed, "multiobject tracking" procedure, many dots are presented on the computer screen at once. Several of the dots begin to flash on and off,and then cease flashing. All of the dots then begin to wander around the screen for a period of time and then come to a halt. The task is to identify the final resting locations of the dots that had been flashing. In this task, it is found that only about 4 dots can be tracked at one time; if a larger number of dots flash, their locations are not correctly tracked. Finally, (f) Fisher (1984) developed a model of consistently mapped search that included a limit in the number of comparisons that could be made at one time. Subjects examined a series of rapidly presented character arrays and searched for the single digit "5." The ability to detect this target depended on both the presentation duration of each array and the number of items in each array. In the mathematical model of performance, based on Erlang's Loss Formula, characters in the visual array enter the information-processing system in parallel for comparison with the target in memory, and if a comparison channel is unavailable at the time of arrival then that particular comparison is not made and the stimulus item is lost from the system. The model provided a very good fit to the data, with the best fit for each individual occurring with 3 to 5 comparison channels.

There are numerous other procedures that yield seemingly similar results. Research is needed to determine the extent to which these diverse procedures are tapping the same mental faculty, and the extent to which their convergent findings are merely coincidental. Until a common mechanism can be unambiguously identified (e.g., some structural or functional limit in the capacity of the focus of attention), we can only be haunted by the integer plus or minus 1. However, the procedures leading to this constant do seem to me to be more numerous even than the number 7 plus or minus 2 that haunted Miller (1956) and could well reflect a limit in what can enter the focus of awareness. The limit may or may not be a discrete integer value and probably applies only to the number of unconnected units or chunks (Miller, 1956), with grouping and chunking processes resulting in a larger effective limit.

TIME LIMIT? The only time limit in attention that is apparent is that reflected in the research on vigilance (e.g., Davies & Parasuraman, 1982). That research suggests that it is impossible to maintain maximal attention on an object indefinitely because the state of alertness wanes (e.g., after $^1/_2$ hour).

The Role of Working Memory in Complex Cognitive Activities

In the present view (cf. Cowan, 1988, 1995), a distinction is to be drawn between the fundamental mechanisms of working memory and the performance on a task. The idea of working memory is simply that the information needed to do a task must be made especially accessible temporarily; often, for a creative task, several pieces of information must be active concurrently so that they can be combined. Thus, I agree with Baars's (1988) view of working memory as a global workspace within which information is integrated rather than held in isolated bundles. However, the mechanisms through which information can become accessible vary widely. Some of the necessary information may be in the focus of attention; some may be in an especially active state, ready to enter the focus as needed; and some may simply have the appropriate contextual coding in long-term memory that allows it to be made available quickly.

Thus, there is no single, separate theoretical entity that I would call working memory; that is a practical, task-oriented label. What are potentially more meaningful in a theoretical sense are the basic mechanisms proposed to underlie this complex system, including activation of memory, the contents of an attentional process, and the contextual organization of memory.

A further important question is what types of information become active or become the focus of attention in complex tasks. My assumption is that multiple types of features become active. However, the focus of attention will prolong this activation for some types of features more than others. Studies of language comprehension in normal and brain-damaged patients clearly show that language is retained in primarily a phonological form only if the language is challenging or complex, in which case the correct syntactic and lexical analysis may not become immediately clear to the listener or reader (e.g., Caplan & Waters, 1990; Martin, 1993). Thus, attention may be focused so as to select, from among the available types of representation, the ones affording the most efficiency and usefulness given a limited attentional capacity.

The Relation of Working Memory to Long-Term Memory and Knowledge

At any moment there is assumed to be a currently active subset of long-term memory, and the focus of attention is assumed to be a subset of that activated information. This attentional focus and the activated memory both play a role in working memory. Both Ericsson and Delaney (Chapter 8, this volume) and Cowan (1995, chapter 4) suggested that items in long-term memory that are strongly associated with the subject's current situation can be accessed easily and function much as if they were held in an activated form. Cowan (1995, chapter 4) called this use of long-term memory as if it were part of short-term memory "virtual" (as opposed to actual) short-term memory, and

Ericsson and Kintsch (1995) called it "long-term working memory." For complex tasks, Ericsson and Delaney (Chapter 8, this volume) describe the processes of long-term working memory in much more detail. However, long-term memory plays a role even in ordinary memory span tasks. Spans are higher for word stimuli than for nonsense word stimuli, the only difference between them being a long-term memory representation available for the words (Hulme, Maughan, & Brown, 1991). Cowan's (1995) "virtual short-term memory" concept was developed to account for such findings and for list-recall results that mimic short-term memory results, but over periods of time so long that they could not actually involve short-term memory mechanisms (e.g., Bjork & Whitten, 1974).

Despite this proposed use of long-term memory, the concept of memory activation is included in the present model given the existing evidence (e.g., Figure 3.6). Also, for success in some tasks, it may be necessary for several pieces of information to be activated concurrently, not just available in long-term memory. For example, it may well be that the structure of an English sentence can be adequately understood only if the subject and verb are active in memory at the same time.

Finally, there is one important qualification of the statement that working memory contains activated elements of long-term memory. Most stimulus situations in life include novel combinations of familiar features. In memory the elements are activated independently, but the particular links between those elements are often novel. The current combination of elements may, however, be stored as a new long-term memory trace. Declarative memories are said to be encoded only with the presence of attention, whereas procedural memories might be encoded more automatically, provided that sufficient attention is devoted to the task to allow the relevant stimulus features to be processed (Cowan, 1995).

The Relationship of Working Memory to Attention and Consciousness

I assume that, in neurally intact individuals, the information in the focus of attention is the same information that the person is aware of, which is also the same information to which a central capacity limit applies (Cowan, 1988, 1995). (See my definition of attention in the introduction.) One type of evidence for this assumption is the close relation between measures of attention shifting to an event in a task-irrelevant channel and subsequent measures of memory of that event (Wood & Cowan, 1995; see Figure 3.3). Activated memory outside of awareness also can count as part of working memory because that information is more readily available to influence task performance than information that is not activated. Presumably, most of the 7 ± 2 chunks of information that people can remember (or whatever the actual limit is) are activated but still are outside the focus of attention at any one moment, given the more severely limited capacity of awareness observed in a number of situations (see above). Covert rehearsal may serve to reactivate information by recirculating it through the focus of attention.

The Biological Implementation of Working Memory

Two main aspects of the biological implementation of the present model merit consideration. First, some psychophysiological research is relevant to the functioning of working memory. Second, other research on lesions and neuroimaging is relevant to the neuroanatomical representation of working memory. These aspects will be addressed in turn.

Psychophysiological Investigations of Working Memory Function

The model of working memory of Cowan (1988) was used by Cowan, Winkler, Teder, and Näätänen (1993) to generate hypotheses about event-related potentials (ERPs). The topic of research was the mismatch negativity (MMN) and its interpretation. The MMN is obtained by repeating an auditory stimulus presentation and then changing the stimulus in some discriminable way. The changed or "deviant" stimuli produce an average waveform that at a certain point becomes more negative than the average waveform produced by the standard stimuli (for a review see Ritter, Deacon, Gomes, Javitt, & Vaughan, 1995). This negative-going difference wave can be obtained for any discriminable acoustic change, even if the subject carries out a different task and ignores the auditory stimuli when they are presented. However, the MMN occurs only if the standard and deviant sounds occur within about 10 s of one another (Mäntysalo & Näätänen, 1987; Sams, Hari, Rif, & Knuutila, 1993). The interpretation has been that a short-lived sensory representation of the standard is compared to the deviant, and the MMN results from a process in which these stimuli are compared and found to differ. This process cannot take place if the sensory representation of the standard has decayed away before the deviant is presented.

One anomaly of this account is that it appeared from various studies that multiple presentations of the standard were necessary before a deviant would produce the MMN. That should not be the case if the standard representation is purely a temporary representation of the last stimulus and is not influenced by the past history of stimulation. A more sophisticated concept of the sensory representation was needed.

Such a concept can be derived from the model in Figure 3.1. According to this model, a repeated presentation of the standard tone should produce a long-term memory representation of that tone. The long-term representation then could be in an active state (likely if the last standard tone occurred about 10 s ago or less) or an inactive state (likely if the last standard tone occurred much prior to that). It was proposed that the MMN would occur only if (a) an adequate representation of the standard tone existed, and (b) that representation was in an active state when the deviant tone was presented.

Cowan et al. (1993) examined this account by presenting trains of nine tones with 610 ms onset-to-onset intervals within a train and silent periods of 11 to 15 s between trains. The tones in each train were all the same except that one deviant could occur, in Serial Positions 1, 2, 4, 6, or 8 within a train. The MMN was examined by comparing ERPs for the deviants in a particular serial position to ERPs for standard tones in the same serial position within trains that had no deviants up to that serial position. There were two kinds of sequences: "constant-standard" sequences in which the standard tone frequency was the same in each train, and "roving-standard" sequences in which a new standard tone was used in each train. In either case, if a deviant occurred it was 7/6 the frequency of the standard tones in its train.

The results beautifully matched the theory. Several standards are needed for the construction of a representation of that tone. In the roving-standard condition, that representation had to be built up for each train anew. Accordingly, there was no MMN for deviants in Serial Positions 1 or 2 within the train, but there was an MMN for deviants later in the train. In the constant-standard condition, however, a long-term memory representation of the standard tone could be built up across trains. However, the intertrain interval would render that representation inactive, so a single reminder presentation would be needed to reactivate the standard tone representation. As this line of reasoning predicts, the MMN did not occur with the deviant in Serial Position 1, but it did occur with the deviant in Serial Position 2 or later. (For a Position 2 deviant, the Position 1 standard was assumed to reactivate the representation.) The magnitude of the MMN in each condition is shown in Figure 3.8. It shows that a deviant in the second serial position elicited an MMN in the constant-standard condition (because the first serial position standard could reactivate the previous standard representation) but not in the roving-standard condition (because the standard was new with the first serial position stimulus). This research thus helps to confirm the theory's emphasis on the concept of the activation of long-term memory as a factor in working memory.

Another main emphasis of the theory is on the habituation and dishabituation of the orienting response as mechanisms regulating the entry of information into the focus of attention automatically. Some of the psychophysiological research on habituation of the orienting response helped to provide the basis of the theory in the first place, showing that the orienting response occurs when there is a change in the basic properties of the stimuli or the appearance of an especially significant stimulus (e.g., Gati & Ben-Shakhar, 1990; Öhman, 1979; Waters et al., 1977).

One aspect of the theoretical rationale is especially important in understanding a broad range of behavioral and psychophysiological evidence on what causes an orienting response. Features of the stimuli that have been extracted from the environment can be added to the neural model that is used in a comparison with subsequent stimuli. However, features that have not been extracted obviously cannot contribute. More features will be

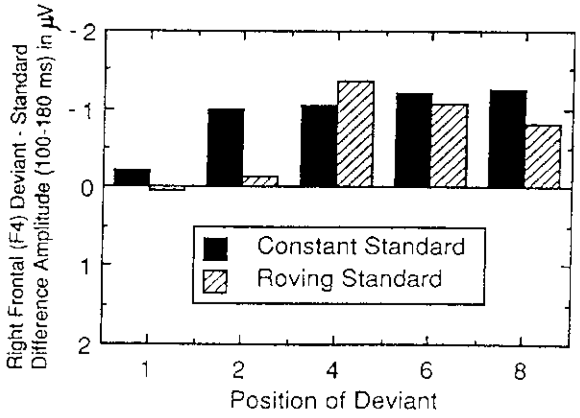
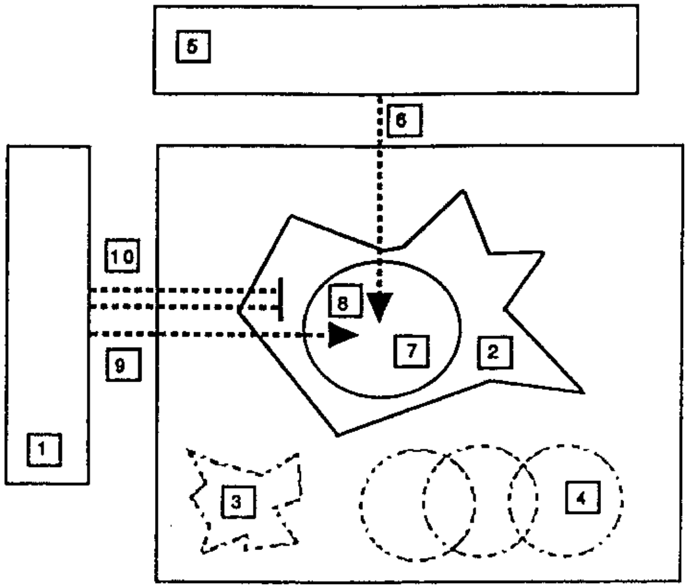


Figure 3.8. Magnitude of the mismatch negativity response in auditory event-related potentials, for constant- and roving-standard conditions with the deviant at various serial positions within a train. Reprinted by permission of the American Psychological Association from Cowan, Winkler, Teder, and Näätänen (1993, p. 915, Figure 3). Copyright © 1993 by the American Psychological Association.

extracted for attended stimuli than for unattended stimuli, and therefore the orienting processes will differ. For example, if an unattended speech channel switches from a discussion of beans to a discussion of flowers, with no change in the physical properties of the channel, no orienting response would be expected. If the channel is attended, the semantic change would be encoded and an orienting response would be expected. The orienting-response literature appears to bear out this difference.

Neuroanatomical Investigations of Working Memory

Cowan (1995, pp. 249–272) ventured to suggest what areas or systems of the brain appear to be involved in each aspect of the model from Cowan (1988), based on literature that is largely beyond the scope of the present chapter. A diagram briefly summarizes the proposals (Figure 3.9). Most of them should be taken as hypotheses to help guide further investigations. Such conjectures may be important because there have been few attempts to synthesize neurological findings related to a broad spectrum of information-processing mechanisms. (1) In electrophysiological and neuroimaging studies related to *sensory memory*, activity in modality-specific cortical areas has been observed; (2) *activated memory* is assumed to be mediated by diverse areas of association cortex, as they appear to become active not only with sensory input but also with the mental imagery and retrieval (e.g., temporal association cortex for auditory imagery); (3) *automatic long-term storage* was tentatively assumed to conform to the proceduralist view in which the brain areas



	Cognitive Component	Likely Neural	Possible Neural
1	Brief sensory store	Sensory cortex	Cell assembl. activ.
2	Memory activation	Association cortex	Synaptic plasticity
3	Automatic storage	Diffuse cortical	Hebbian circuits
4	Storage with attention	Hippocampus	Septum, Frontal lobes
5	Central executive	Prefrontal cortex	
6	Attentional innervation	Thalamus/ARAS Support	
7	Focus of attention	Parietal lobes	Frontal lobes/Thalamus
8	Entry into focus	Thalamus	Synch. Oscillations
9	Orienting of Attention	Locus C./Hippoc./Parietal	Frontal lobes
10	Habituation of Orienting	Hippocampus	Frontal lobes

Figure 3.9. Suggestions of areas in the brain underlying various components of the model. Reprinted by permission of Oxford University Press from Cowan (1995, p. 250, Figure 8.1). Copyright © 1995 by Oxford University Press.

active in the original processing of a stimulus or event (often distributed across many areas of the brain) are also involved in the storage of the information related to the event; (4) *attention-related long-term storage* was assumed to take place also in diverse cortical areas, but with the additional critical involvement of the hippocampus and surrounding brain areas, resulting in a much more accessible memory record; (5) the *central executive* was strongly linked to the prefrontal cortex on the basis of neurological deficits and neuroimaging results; (6) based largely on the neuroanatomical organization of the brain, it was suggested that the *innervation or source of attentional activity* flows through the thalamus with the support of the ascending reticular activating system (ARAS); (7) it was suggested that the *focus of attention or awareness* critically involves areas of the parietal lobes, which receive signals from the frontal lobes, the ARAS, and the various sensory systems (the basis for this assumption being that deficits in awareness arise primarily from parietal damage, as opposed to deficits in control, which arise primarily from frontal damage); (8) *entry of information into the focus of attention or awareness* was said to occur via the thalamus, with the possibility that there are synchronous oscillations of various areas representing features of the attended object; and (9) the *orienting of attention* was said to involve the locus ceruleus among other areas, as shown in physiological research, whereas (10) its *habituation* requires the formation of a neural model built up with the help of various cortical areas and hippocampal involvement; with hippocampal damage, habituation of orienting appears to be impaired.

Below I will discuss just two especially critical assumptions in further detail: those related to memory activation (Conjecture 2) and the focus of attention and awareness (Conjecture 7).

MEMORY ACTIVATION. The "memory activation" concept subsumes activities for various modalities and codes, in various parts of the brain. The point of discussing them together under the term *activation* is to acknowledge that these types of activity may be fundamentally similar in their mode of operation in the brain. For example, Cowan (1988) discussed similarities in the time scales for a literal sensory afterimage of the stimulus (about 200 to 300 ms) and for a secondary, more processed sensory representation (about 20 to 30 s) in the auditory, visual, and tactile modalities, and potentially in all modalities.

The areas of the brain involved in perceiving stimuli are assumed to be the same as, or adjacent to, the areas containing the activated representations of the stimuli. The link between perception and activated memory already seems clear in the case of auditory sensory information. Thus, magnetoencephalographic studies indicate temporary changes in the auditory cortex concurrent with auditory sensory memory and its decay (Hari, Kaila, Katila, Tuomisto, & Varpula, 1982; Lü, Williamson, & Kaufman, 1992; Sams et al., 1993), and damage to the temporal cortex has been shown to produce a loss of auditory information in monkeys that is much more rapid and severe than normal (Colombo, D'Amato, Rodman, & Gross, 1990). Perceived, remembered, or imagined features of a particular stimulus all may activate neurons in similar areas of the brain (e.g., Farah, 1988), although truly sensory features in working memory are not fully interchangeable with remembered or imagined features (e.g., Keller et al., 1995). In recent work, there are leads helping to distinguish automatic activation from working-memory processes that require attention (e.g., Miller & Desimone, 1994).

THE FOCUS OF ATTENTION AND AWARENESS. One of the key issues of our time is the neural representation of the focus of attention and awareness. There have been a number of proposals about this, and it is clear that the attentional system encompasses much of the brain (including frontal, parietal, and thalamic areas among others). However, a more specific and difficult challenge will be to differentiate this system into subsystems; for example, those subserving (a) central executive functioning, (b) attention switching and attention maintenance, and (c) the focus of attention or seat of awareness.

One popular view is that the frontal lobes mediate consciousness. However, I do not see how to reconcile it with the neurological evidence that frontal brain damage does not generally result in a loss of awareness, but rather in an impairment of central executive functions such as the maintenance of attention and the instigation, control, organization, planning, and execution of complex activities directed toward a goal. Cowan (1995, pp. 257–261) reviewed evidence that deficits of awareness are more characteristic of inferior parietal lobe damage (e.g., see Schacter, 1989). The parietal areas are typically involved, for example, in cases of anosognosia, in which the patient may be partly paralyzed but denies having any deficit, and in cases of unilateral neglect. Its involvement in attention mechanisms has been studied mostly with visual stimuli, but is not limited to that modality. The inferior parietal lobe and surrounding areas also are qualified for the "focus of attention" role in that they (in particular, Area 7 along with the adjoining Area Tpt of the temporal lobe) are thought by some researchers to be involved in the integration of data from all of the senses together, and perhaps uniquely so (e.g., Hyvärinen, 1982). Thus, there is reason to suspect that these areas (and parts of the thalamus that innervate them) may play a large role in the seat of awareness as opposed to its control.

Research on the parietal lobes has helped to make the point that not all processing is accompanied by conscious awareness of the processed information. In one striking example, Volpe, Ledoux, and Gazzaniga (1979) studied the ability of four patients with right parietal lobe tumors to (a) identify objects in the left and right hemifields and (b) compare objects across hemifields. These patients could make same–different comparisons between the objects in the two hemifields very well (88–100% correct, mean = 93%), but still they were generally unable to identify the objects in the left hemifield (0–48% correct, mean = 18%). For example, one subject correctly responded "different," and when then asked, "what exactly?" responded, "a comb and I don't know what the other was." The parietal lobe damage appears to have affected awareness of the objects without having affected the processing needed for the same–different comparison.

Attention and awareness generally coincide in the normal individual, but they are not logically inseparable. If the left and right parietal lobes are prevented from exchanging signals, as in split-brain individuals (who do not have intact corpus collosums), awareness should be divided in two; and that is the case. However, if attention were the same as awareness, then the attentional capacity also should be divided in two, or at least limited separately in each hemisphere. That, however, is not what is found. Instead, a central capacity limit remains even though awareness is divided into two separated entities. Holtzman and Gazzaniga (1982) found that perceptual tasks in the left and right hemispheres of split-brain patients share processing capacity even though the hemispheres are unaware of each other's tasks. This suggests that attentional capacity and awareness are not logically the same thing, though they work as an integral system normally. There may be a system involving the frontal lobes and thalamic centers that innervates the parietal lobes and determines attentional capacity, whereas the parietal lobe itself is more heavily involved in awareness.

It is important to ask what impact neuroscientific research could have on the present theoretical view. There is a lot of potentially relevant, though as yet indecisive, recent research on the frontal and parietal lobes (e.g., see Beardsley, 1997; Buckner, 1996). Such neuroimaging research theoretically could result in a confirmation, clarification, modification, or disconfirmation of the view espoused by Cowan (1995). A neuroimaging result that simply indicated a dissociation not present in Cowan (1995) is capable of clarifying or modifying the functional theoretical view, but not disconfirming it. For example, the finding that different areas of the prefrontal cortex are involved in different working memory tasks, or that attention and awareness can be dissociated neurally, would not disconfirm the theory, but it might refine it. What would be injurious to the model would be an organization or *alignment* of processes different from the expected one. For example, according to the model, the central executive does not maintain its own duplicate representation of the items in working memory, but instead causes the relevant representations (which exist elsewhere in the brain) to be in the focus of awareness. If it were found that the central executive actually has its own duplicate representation, that would go against the theory. Assuming that certain areas of the prefrontal areas correspond to central executive processes, it should not be found that these areas remain active in a successfully completed working memory trial in the absence of any involvement of other areas in the brain that presumably would contain the memory representation. Moreover, in cases of unattended information held temporarily in a sensory form, the sensory areas but not the frontal lobes should be

involved in the task (though the frontal lobes might be busy with other tasks). Thus, the biological implementation is of potential relevance to the cognitive level of modeling.

Concluding Observations

The concept of working memory is central to cognitive psychology because it is assumed to be the vehicle for the retrieval of all information that is needed to carry out a particular task. As such, it is unlikely to be limited to one mechanism; people are likely to use any processing mechanisms at their disposal to come up with the needed information. The use of several mechanisms usually is less taxing than the reliance on any one mechanism. Cowan (1988, 1995) described a processing framework based on the premise that memory activation mechanisms, attentional and executive mechanisms, and long-term retrieval mechanisms all work together in processing to form an effective working memory system. There are still many unknowns, but the modeling framework is designed to organize the knowns and encourage direct tests of some fundamental, unproven assumptions.

REFERENCES

Baars, B. J. (1988). *A cognitive theory of consciousness.* New York: Cambridge University Press.

Baddeley, A. D. (1986). *Working memory.* New York: Oxford University Press.

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behavior, 22,* 88–104.

Banbury, S., & Berry, D. C. (1997). Habituation and dishabituation to speech and office noise. *Journal of Experimental Psychology: Applied, 3,* 181–195.

Beardsley, T. (1997, August). The machinery of thought. *Scientific American, 277,* 78–83.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6,* 173–189.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon Press.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3–18). New York: Wiley.

Buckner, R. L. (1996). Beyond HERA: Contributions of specific prefrontal brain areas to long-term memory retrieval. *Psychonomic Bulletin & Review, 3,* 149–158.

Cantor, J., & Engle, R. W. (1993). Working memory capacity as long-term memory activation: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19,* 1101–1114.

Caplan, D., & Waters, G. S. (1990). Short-term memory and language comprehension: A critical review of the neuropsychological literature. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 337–389). New York: Cambridge University Press.

Chase, W. G. (1977). Does memory scanning involve implicit speech? In S. Dornic (Ed.), *Attention and performance VI* (pp. 607–628). Hillsdale, NJ: Erlbaum.

Clifton, C., & Tash, J. (1973). Effect of syllabic word length on memory-search rate. *Journal of Experimental Psychology, 99,* 231–235.

Colombo, M., D'Amato, M. R., Rodman, H. R., & Gross, C. G. (1990). Auditory association cortex lesions impair auditory short-term memory in monkeys. *Science, 246,* 336–338.

Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology, 55,* 75–84.

Conrad, R., & Hille, B. A. (1958). The decay theory of immediate memory and paced recall. *Canadian Journal of Psychology, 12,* 1–6.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123,* 354–373.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96,* 341–370.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104,* 163–191.

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language, 31,* 668–684.

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York: Oxford University Press.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language, 31,* 1–17.

Cowan, N., Keller, T. A., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing clues to the mechanisms underlying age and word length effects. *Journal of Memory and Language, 33,* 234–250.

Cowan, N., Lichty, W., & Grove, T. R. (1990). Properties of memory for unattended spoken syllables. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16,* 258–269.

Cowan, N., & Saults, J. S. (1995). Memory for speech. In H. Winitz (Ed.), *Human communication and its disorders* (Vol. 4). Timonium, MD: York Press.

Cowan, N., Saults, J. S., & Nugent, L. D. (1997). The role of absolute and relative amounts of time in forgetting within immediate memory: The case of tone pitch comparisons. *Psychonomic Bulletin & Review., 4,* 393–397.

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of the mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 909–921.

Cowan, N., Wood, N. L., & Borne, D. N. (1994). Reconfirmation of the short-term storage concept. *Psychological Science, 5,* 103–106.

Cowan, N., Wood, N. L., Nugent, L. D. & Treisman, M. (1997). There are two word length effects in verbal short-term memory: Opposed effects of duration and complexity. *Psychological Science, 8,* 290–295.

Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D., & Keller, C. V. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General., 127,* 141–160.

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 472–478.

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition, 21,* 142–145.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3,* 255–267.

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance.* New York: Academic Press.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70,* 80–90.

Engle, R. W., Conway, A. R. A., Tuholski, S. W., & Shisler, R. J. (1995). A resource account of inhibition. *Psychological Science, 6,* 122–125.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95,* 307–317.

Fisher, D. L. (1984). Central capacity limits in consistent mapping, visual search tasks: Four channels or more? *Cognitive Psychology, 16,* 449–484.

Gati, I., & Ben-Shakhar, G. (1990). Novelty and significance in orientation and habituation: A feature-matching approach. *Journal of Experimental Psychology: General, 119,* 251–263.

Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science, 4,* 294–298.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351–360.

Gray, J. A., & Wedderburn, A. A. I. (1960). Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology, 12,* 180–184.

Guttentag, R. E. (1984). The mental effort requirement of cumulative rehearsal: A developmental study. *Journal of Experimental Child Psychology, 37,* 92–106.

Halford, G. S., Maybery, M. T., & Bain, J. D. (1988). Set-size effects in primary memory: An age-related capacity limitation? *Memory & Cognition, 16,* 480–487.

Hari, R., Kaila, K., Katila, T., Tuomisto, T., & Varpula, T. (1982). Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: Implications for their neural generation. *Electroencephalography and Clinical Neurophysiology, 54,* 561–569.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). New York: Academic Press.

Hebb, D. O. (1949). *Organization of behavior.* New York: Wiley.

Hirst, W., Spelke, E. S., Reaves, C. C., Caharack, G., & Neisser, U. (1980). Dividing attention without alternation or automaticity. *Journal of Experimental Psychology: General, 109,* 98–117.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences, 9,* 1–66.

Holtzman, J. D., & Gazzaniga, M. S. (1982). Dual task interactions due exclusively to limits in processing resources. *Science, 218,* 1325–1327.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30,* 685–701.

Hyvärinen, J. (1982). *The parietal cortex of monkey and man.* Berlin: Springer-Verlag.
James, W. (1890). *The principles of psychology.* New York: Henry Holt.
Johnston, W. A., & Heinz, S. P. (1978). Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General, 107,* 420–435.
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.
Keller, T. A., Cowan, N., & Saults, J. S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 635–645.
Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1,* 153–161.
Lü, Z.-L., Williamson, S. J., & Kaufman, L. (1992). Physiological measures predict behavioral lifetime of human auditory sensory memory. *Science, 258,* 1668–1670.
Mäntysalo, S., & Näätänen, R. (1987). The duration of a neuronal trace of an auditory stimulus as indicated by event-related potentials. *Biological Psychology, 24,* 183–195.
Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15,* 197–237.
Martin, R. C. (1993). Short-term memory and sentence processing: Evidence from neuropsychology. *Memory & Cognition, 21,* 176–183.
McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review, 39,* 352–370.
Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science, 263,* 520–522.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.
Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11,* 56–60.
Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.
Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay. *Psychonomic Bulletin & Review, 2,* 429–441.
Öhman, A. (1979). The orienting response, attention, and learning: An information-processing perspective. In H. K. Kimmel, E. H. Van Olst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 443–471). Hillsdale, NJ: Erlbaum.
Peterson, L. R. & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198.
Reitman, J. S. (1974). Without surreptitious rehearsal, information in short-term memory decays. *Journal of Verbal Learning and Verbal Behavior, 13,* 365–377.
Ritter, W., Deacon, D., Gomes, H., Javitt, D. C., & Vaughan, H. G., Jr. (1995). The mismatch negativity of event-related potentials as a probe of transient auditory memory: A review. *Ear and Hearing, 16,* 52–67.
Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The human auditory sensory memory trace persists about 10 sec: Neuromagnetic evidence. *Journal of Cognitive Neuroscience, 5,* 363–370.
Schacter, D. L. (1989). On the relation between memory and consciousness: Dissociable interactions and conscious experience. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in Honor of Endel Tulving* (pp. 355–389). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.
Sokolov, E. N. (1963). *Perception and the conditioned reflex.* New York: Pergamon Press.
Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74* (Whole No. 498).
Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654.
Treisman, A. M. (1964). Selective attention in man. *British Medical Bulletin, 20,* 12–16.
Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 331–351.
Volpe, B. T., Ledoux, J. E., & Gazzaniga, M. S. (1979). Information processing of visual stimuli in an "extinguished" field. *Nature, 282,* 722–724.
Waters, W. F., McDonald, D. G., & Koresko, R. L. (1977). Habituation of the orienting response: A gating mechanism subserving selective attention. *Psychophysiology, 14,* 228–236.
Watkins, M. J. (1974). Concept and measurement of primary memory. *Psychological Bulletin, 81,* 695–711.
Watkins, M. J., Watkins, O. C., Craik, F. I. M., & Mazuryk, G. (1973). Effect of nonverbal distraction of short-term storage. *Journal of Experimental Psychology, 101,* 296–300.
Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). New York: Academic Press.
Wingfield, A., & Byrnes, D. L. (1972). Decay of information in short-term memory. *Science, 176,* 690–692.
Wood, N., & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 255–260.
Wood, N. L., Stadler, M. A., & Cowan, N. (1997). Is there implicit memory without attention? A re-examination of task demands in Eich's (1984) procedure. *Memory & Cognition, 25,* 772–779.